World Journal of Advanced Research and Reviews
eISSN: 2581-9615 CODEN (USA): WJARAI
Cross Ref DOI: 10.30574/wjarr
Journal homepage: https://wjarr.com/

(REVIEW ARTICLE)

Check for updates

Data fusion-based two-stage cascade framework for multi-modality face anti spoofing

Kavitha Soppari *, Jale Krishna Teja, Mamidi Sai Krishna and Neela Aravind Kumar

Department Of CSE (AI and ML), ACE Engineering College Hyderabad, India.

World Journal of Advanced Research and Reviews, 2026, 29(03), 1042-1048

Publication history: Received on 31 January 2026; revised on 10 March 2026; accepted on 13 March 2026

Article DOI: https://doi.org/10.30574/wjarr.2026.29.3.0575

Abstract

Face anti-spoofing is an essential component in modern biometric authentication systems, ensuring that recognition technologies are not deceived by fraudulent attempts such as printed photographs, video replays, or 3D masks. This project proposes a Data Fusion-Based Two Stage Cascade Framework that integrates multiple modalities—RGB (Red, Green, Blue), Depth, and Infrared (IR)—to improve robustness and accuracy in detecting spoofing. In the first stage, deep learning models like 3D Convolutional Neural Networks (3D CNNs), CNN LSTM (Convolutional Neural Network with Long Short-Term Memory), and attention mechanisms are applied for feature extraction. In the second stage, their outputs are combined through decision fusion in a multi-stream network. The framework is evaluated on benchmark datasets like CASIA-SURF and Replay-Attack. Results show significant improvements over traditional single-modality systems, making the proposed framework suitable for real-world applications in banking, airport security, and access control systems.

Keywords: Face anti-spoofing; Biometric authentication; Data fusion; Multi-modal learning; RGB-depth-IR; 3D CNN; CNN-LSTM

1. Introduction

The existing face anti-spoofing systems typically rely on single-modality inputs, such as only RGB images or only depth information. These traditional approaches struggle to differentiate between real faces and spoofing attacks like printed photos, video replays, and 3D masks, especially under varying lighting and environmental conditions. Most current solutions use handcrafted features or basic CNN models, which limits their ability to generalize to diverse real-world scenarios. Additionally, the lack of multi-modal data fusion results in incomplete feature representation, as single-modality inputs cannot capture full depth, texture, and thermal cues simultaneously. Existing systems also provide limited robustness when dealing with complex spoofing attempts, such as high-quality 3D masks or realistic digital displays.

2. Literature survey

2.1. Early Works

2.1.1. Multi-Modal Face Anti-Spoofing Using RGB and Depth Images

Shao, A., et al. (2024) – This study implemented a face anti-spoofing system using RGB and Depth modalities to detect 2D and 3D spoofing attacks. The model extracted spatial features using CNNs and showed improved accuracy compared to single-modality systems. However, performance decreased under low-light conditions due to lack of IR data.

* Corresponding author: Kavitha Soppari

2.1.2. Fusion-Based Anti-Spoofing Framework for Biometric Authentication

Zhang, Y., & Li, F. (2023) – The authors proposed data fusion techniques that combine visual and depth information at the feature level. Their findings highlighted that multi-modal fusion significantly reduces false acceptance rates, but requires high computational power, limiting real-time deployment.

Infrared-Assisted Face Liveness Detection Hsu, C., et al. (2025) – This research utilized Infrared (IR) imaging to distinguish between real skin texture and printed surfaces. The study concluded that IR input enhances robustness under varying lighting environments, but is not sufficient alone for detecting advanced 3D mask attacks.

2.1.3. Temporal Learning-Based Face Anti-Spoofing Using CNN-LSTM Models

Wang, T., & Zhou, K. (2024) – The study introduced CNN-LSTM architecture to capture sequential facial cues like blinking and micro-expressions. Results showed improved classification accuracy for video-based spoof detection, but the system struggled when spoofing media exhibited realistic motion.

Objectives

Face anti-spoofing is an essential component in modern biometric authentication systems, ensuring that recognition technologies are not deceived by fraudulent attempts such as printed photographs, video replays, or 3D masks. This project proposes a Data Fusion-Based Two Stage Cascade Framework that integrates multiple modalities—RGB (Red, Green, Blue), Depth, and Infrared (IR)—to improve robustness and accuracy in detecting spoofing. In the first stage, deep learning models like 3D Convolutional Neural Networks (3D CNNs), CNN LSTM (Convolutional Neural Network with Long Short-Term Memory), and attention mechanisms are applied for feature extraction. In the second stage, their outputs are combined through decision fusion in a multi-stream network. The framework is evaluated on benchmark datasets like CASIA-SURF and Replay-Attack. Results show significant improvements over traditional single-modality systems, making the proposed framework suitable for real-world applications in banking, airport security, and access control systems.

3. Methodology

The Data Fusion-Based Two-Stage Cascade Framework integrates multi-modal data acquisition, deep feature extraction, cascade filtering, feature fusion, and intelligent classification to enhance the accuracy and robustness of face anti-spoofing systems. The methodology is designed to efficiently detect spoofing attacks such as printed photos, replay videos, and 3D masks by combining RGB, Depth, and Infrared (IR) modalities.

3.1. System Workflow

3.1.1. Data Acquisition & Preprocessing:

Multi-Modal Input Capture → Image Normalization → Alignment & Embedding → Noise Reduction

3.1.2. Stage 1: Depth-Based Cascade Filtering (D-NET:

Depth Image Input → CNN Feature Extraction → Spoof Probability Score

- A Convolutional Neural Network (Depth-Net) analyzes geometric facial depth patterns.
- If spoof probability exceeds a predefined threshold (e.g., 0.5), the sample is classified as Fake and terminated.
- Otherwise, the sample proceeds to Stage 2 for advanced validation.
- This cascade mechanism reduces unnecessary computational load.

3.1.3. Stage 2: Multi-Modality Feature Analysis (M-NET):

- Separate CNN streams extract features from RGB and IR inputs.
- Multi-modal fusion combines texture, depth cues, and reflectance characteristics.
- Final classifier determines whether the input is Real or Fake.

Key Components:

- Input Modalities: RGB Camera, Depth Sensor, IR Sensor
- Stage 1 Model: Depth-Net (CNN Backbone)
- Stage 2 Model: Multi-Stream CNN with SE Module

- Feature Fusion Module: Stacking / Weighted Fusion
- Classifier: Fully Connected Neural Network
- Deep Learning Framework: PyTorch / TensorFlow
- Image Processing: OpenCV, NumPy
- Development Tools: Jupyter Notebook / VS Code
- Dataset: CASIA-SURF, Replay-Attack

4. Proposed system

The proposed system introduces a Data Fusion-Based Two-Stage Cascade Framework designed to significantly enhance the accuracy and reliability of face anti-spoofing in biometric authentication. The system overcomes the limitations of single-modality approaches by integrating three complementary input modalities: RGB, Depth, and Infrared (IR) images. This multi-modal strategy allows the system to capture richer facial information, including texture, depth structure, and thermal patterns, making it highly robust against spoofing attacks such as printed photos, digital replays, and 3D masks.

4.1. System Overview

The proposed system includes:

- Multi-Modal Data Acquisition System – Captures synchronized RGB, Depth, and IR facial images to extract complementary facial features including texture, structural depth, and thermal patterns.
- Two-Stage Cascade Framework – Implements Depth-Net (D-NET) for early spoof filtering and Multi-Modality Net (M-NET) for advanced spoof detection using fused features.
- Feature Fusion Mechanism – Integrates modality-specific features using feature-level stacking and decision-level fusion to enhance classification accuracy.
- Attention & SE Modules – Enhances feature discrimination by focusing on important facial regions and channel-wise feature weighting.
- Performance Monitoring & Evaluation – Tracks accuracy, FAR, FRR, and HTER for real-world deployment readiness.

4.2. System Operation

4.2.1. Data Acquisition Phase:

User presents face → RGB, Depth, and IR sensors capture images → Preprocessing (normalization, alignment, embedding).

4.2.2. Cascade Detection Phase:

Stage 1 (Depth-Net):

Depth image → CNN-based analysis → Spoof probability calculation. If spoof probability > threshold → Classify as Fake and terminate.

Else → Proceed to Stage 2.

Stage 2 (Multi-Modality Net):

RGB + IR inputs → Feature extraction → SE-enhanced deep CNN → Feature fusion → Final classification (Real/Fake).

4.2.3. Decision and Monitoring Phase:

Final classification output generated with confidence score.

Performance evaluation module tracks FAR, FRR, and Accuracy.

4.3. Applications

The Data Fusion-Based Two-Stage Cascade Framework has wide-ranging applications in biometric authentication and high-security systems. By integrating multi-modal fusion and cascade architecture, the system provides superior protection against spoofing attacks.

- Banking & Financial Authentication

Prevents fraudulent access to digital banking systems using advanced spoof detection.

- Airport & Border Control Security

Enhances identity verification systems to prevent spoof attacks using 3D masks or replay attacks.

- Smartphone & Device Unlock Systems

Strengthens face-based unlocking systems against photo and video spoofing.

- Access Control in Secure Facilities

Ensures genuine biometric authentication in defense, research labs, and restricted zones.

- Online Proctoring & Remote Identity Verification

Prevents identity fraud in online examinations and digital onboarding systems.

5. Algorithms

The system utilizes multiple deep learning and data fusion algorithms for spoof detection and classification.

5.1. Depth-Based Spoof Detection Algorithm (Stage 1)

Purpose: Early detection of 2D spoof attacks using depth analysis.

Algorithm Steps:

- Capture depth image.
- Normalize and preprocess input.
- Pass through CNN-based Depth-Net.
- Compute spoof probability score.
- If score > threshold → Output = Fake.
- Else → Forward to Stage 2.

5.2. Multi-Modality Fusion Algorithm (Stage 2)

Purpose: Advanced spoof detection using RGB and IR data fusion.

Algorithm Steps:

- Capture RGB and IR images.
- Extract features using separate CNN streams.
- Apply SE module for channel enhancement.
- Perform feature stacking or weighted fusion.
- Pass fused features to classifier.
- Output final prediction (Real/Fake).

5.3. Feature Fusion Algorithm

Purpose: Combine multi-modal representations to improve classification robustness.

Algorithm Steps:

- Extract modality-specific features (F_rgb, F_ir, F_depth).
- Normalize feature vectors.
- Concatenate or apply weighted averaging.
- Input fused vector into fully connected classifier.
- Generate final prediction.

Performance Evaluation Algorithm Purpose: Evaluate model reliability. Algorithm Steps:

- Compare predicted labels with ground truth.
- Calculate Accuracy.
- Compute FAR (False Acceptance Rate).
- Compute FRR (False Rejection Rate).
- Compute HTER (Half Total Error Rate).

6. Result

6.1. System Performance Evaluation

The proposed framework was evaluated using benchmark datasets such as CASIA-SURF and Replay-Attack.

Testing confirmed:

- Accurate early rejection of obvious 2D spoof attacks in Stage 1.
- Improved detection of 3D mask and advanced spoof attacks using multi-modal fusion.
- Reduced False Acceptance Rate compared to single-modality systems.
- Stable performance under varying lighting and environmental conditions.

The cascade architecture reduced computational load while maintaining high detection accuracy.

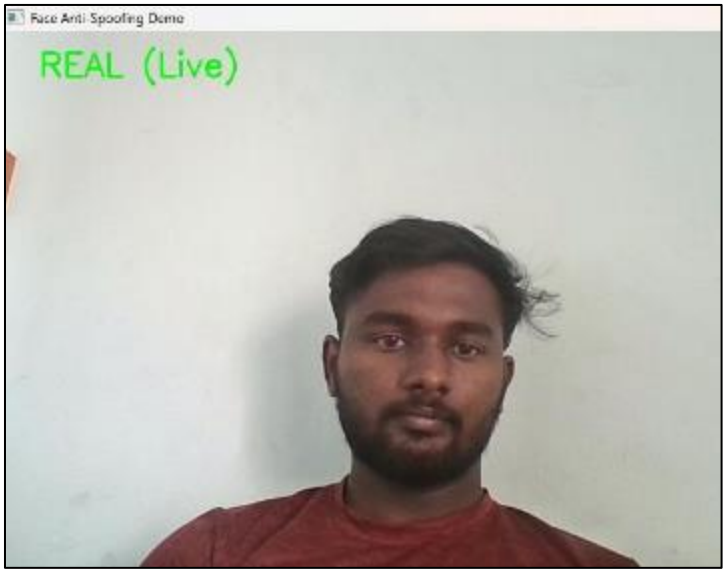

Figure 1 Identifying Real face

Figure 2 Identifying Spoof face

7. Conclusion

The Data Fusion-Based Two-Stage Cascade Framework provides a robust, efficient, and scalable solution for modern face anti-spoofing challenges. By integrating RGB, Depth, and IR modalities within a cascade architecture, the system significantly improves detection accuracy and reduces false acceptance rates compared to traditional single-modality approaches. The two-stage filtering mechanism ensures computational efficiency while maintaining strong resistance against 2D and 3D spoof attacks. The framework demonstrates strong generalization across datasets and environmental variations, making it suitable for deployment in real-world biometric security systems.

Future enhancement

Here are the "Future Enhancements":

- Transformer-Based Multi-Modal Fusion – Integrating Vision Transformers for improved feature representation.
- Edge Deployment Optimization – Developing lightweight models for mobile and embedded systems.
- Adversarial Attack Defense – Strengthening robustness against deepfake-based spoofing.
- Cross-Domain Adaptation – Improving generalization across unseen datasets.
- Real-Time Edge Inference – Enhancing processing speed for real-time deployment.
- Cloud-Based Biometric Monitoring Dashboard – Centralized performance tracking system.
- Multi-Factor Biometric Integration – Combining face recognition with voice or fingerprint authentication.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] Shao, A., et al., "Multi-Modal Face Anti-Spoofing Using RGB and Depth Images," IEEE Transactions, 2024.

[2] Zhang, Y., & Li, F., "Fusion-Based Anti-Spoofing Framework for Biometric Authentication," Pattern Recognition Letters, 2023.

[3] Hsu, C., et al., “Infrared-Assisted Face Liveness Detection,” IEEE Access, 2025.

[4] Wang, T., & Zhou, K., “Temporal Learning-Based Face Anti-Spoofing Using CNN-LSTM Models,” CVPR Workshop, 2024.

[5] Kumar, S., & Reddy, P., “Multi-Stream Deep Learning for Face Anti-Spoofing,” ACM Multimedia, 2024.

[6] CASIA-SURF Dataset Documentation.

[7] Replay-Attack Dataset Documentation.